IMS Collections
???
Vol. 0 (0000) 1–14

arXiv: math.ST/1107.0417

On Low-Dimensional Projections of High-Dimensional Distributions

Lutz Dümbgen[1] **and Perla Zerial**[2]

University of Bern and Technical University of Dresden

Abstract: Let P be a probability distribution on q-dimensional space. The so-called Diaconis-Freedman effect means that for a fixed dimension $d << q$, most d-dimensional projections of P look like a scale mixture of spherically symmetric Gaussian distributions. The present paper provides necessary and sufficient conditions for this phenomenon in a suitable asymptotic framework with increasing dimension q. It turns out that the conditions formulated by Diaconis and Freedman (1984) are not only sufficient but necessary as well. Moreover, letting $\widehat{P}$ be the empirical distribution of n independent random vectors with distribution P, we investigate the behavior of the empirical process $\sqrt{n}(\widehat{P} - P)$ under random projections, conditional on $\widehat{P}$.

1. Introduction

A standard method of exploring high-dimensional datasets is to examine various low-dimensional projections thereof. In fact, many statistical procedures are based explicitly or implicitly on a "projection pursuit", cf. [8]. As shown by Diaconis and Freedman [4], under weak regularity conditions on a distribution $P = P^{(q)}$ on $\mathbb{R}^q$, "most" d-dimensional orthonormal projections of P are similar (in the weak topology) to a mixture of centered, spherically symmetric Gaussian distribution on $\mathbb{R}^d$ if q tends to infinity while d is fixed. A graphical demonstration of this disconcerting phenomenon is given by [3]. Precise quantitative analyses are provided by [9, 10] for situations where most projections are approximately Gaussian. The present paper provides further insight into the general phenomenon. We extend the results of [4] in two directions.

Section 2 gives necessary and sufficient conditions on the sequence $(P^{(q)})_{q \geq d}$ such that "most" d-dimensional projections of P are similar to some distribution Q on $\mathbb{R}^d$. It turns out that these conditions are essentially the conditions of [4]. The novelty here is necessity. The limit distribution Q is automatically a mixture of centered, spherically symmetric Gaussian distributions. The family of such measures arises in [5] in a somewhat different context.

More precisely, let $\Gamma = \Gamma^{(q)}$ be uniformly distributed on the set of column-wise orthonormal matrices in $\mathbb{R}^{q \times d}$ (cf. Section 4.2). Defining

$$\gamma^\top P \ := \ \mathcal{L}_{X \sim P}(\gamma^\top X)$$

for $\gamma \in \mathbb{R}^{d \times q}$, we investigate under what conditions the random distribution $\Gamma^\top P$ converges weakly in probability to an arbitrary fixed distribution Q as $q \to \infty$, while d is fixed.

[1] University of Bern, Department of Mathematics and Statistics, Alpeneggstrasse 22, CH-3012 Bern, Switzerland
[2] Technical University of Dresden
AMS 2000 subject classifications: 62E20, 62G20, 62H99

In Section 3 we study the relationship between $P = P^{(q)}$ and the empirical distribution $\widehat{P} = \widehat{P}^{(q,n)}$ of n independent random vectors with distribution P, also independent from the projection matrix $\Gamma = \Gamma^{(q)}$. Suppose that the distributions $P^{(q)}$ satisfy the conditions of Section 2. Then the random distributions $\widehat{P}^{(q,n)}$ satisfy these conditions, too, as q and n tend to infinity. Furthermore, the standardized empirical measure $n^{1/2}\bigl(\Gamma^\top \widehat{P} - \Gamma^\top P\bigr)$ satisfies a conditional Central Limit Theorem *given the data* $\widehat{P}$.

Proofs are deferred to Section 4. The main ingredients are Poincaré's [11] Lemma and a method invented by Hoeffding [7] in order to prove weak convergence of conditional distributions. Further we utilize standard results from weak convergence and empirical process theory.

2. The Diaconis-Freedman Effect

Let us first settle some terminology. A random distribution $\widehat{Q}$ on a separable metric space $(\mathbb{M}, \rho)$ is a mapping from some probability space into the set of Borel probability measures on $\mathbb{M}$ such that $\int f \, d\widehat{Q}$ is measurable for any function $f \in \mathcal{C}_b(\mathbb{M})$, the space of bounded, continuous functions on $\mathbb{M}$. We say that a sequence $(\widehat{Q}_k)_k$ of random distributions on $\mathbb{M}$ converges weakly in probability to some fixed distribution Q if for each $f \in \mathcal{C}_b(\mathbb{M})$,

$$\int f \, d\widehat{Q}_k \ \to_p \ \int f \, dQ \quad \text{as } k \to \infty.$$

In symbols, $\widehat{Q}_k \to_{w,p} Q$ as $k \to \infty$. Standard approximation arguments (e.g. as in [14], Section 1.12) show that $(\widehat{Q}_k)_k$ converges in probability to Q if, and only if,

$$D_{\mathrm{BL}}(\widehat{Q}_k, Q) := \sup_{f \in \mathcal{F}_{\mathrm{BL}}} \Bigl| \int f \, d\widehat{Q}_k - \int f \, dQ \Bigr| \ \to_p \ 0 \quad (k \to \infty),$$

where $\mathcal{F}_{\mathrm{BL}}$ stands for the class of functions $f : \mathbb{M} \to [-1, 1]$ such that $|f(x) - f(y)| \le \rho(x, y)$ for all $x, y \in \mathbb{M}$.

Now we can state the first result. Here and throughout, $\|\cdot\|$ denotes Euclidean norm and $\mathcal{N}_{d,v}$ stands for the Gaussian distribution on $\mathbb{R}^d$ with mean vector 0 and covariance matrix vI_d.

Theorem 2.1. *The following two assertions on the sequence* $(P^{(q)})_{q \ge d}$ *are equivalent:*

(A1) *There exists a probability measure Q on $\mathbb{R}^d$ such that*

$$\Gamma^\top P \ \to_{w,p} \ Q \quad \textit{as } q \to \infty.$$

(A2) *If $X = X^{(q)}, \tilde{X} = \tilde{X}^{(q)}$ are independent random vectors with distribution P, then*

$$\mathcal{L}(\|X\|^2/q) \ \to_w \ R \quad \textit{and} \quad X^\top \tilde{X}/q \ \to_p \ 0 \quad \textit{as } q \to \infty$$

for some probability measure R on $[0, \infty)$.

The limit distribution Q in (A1) is a normal mixture, precisely,

$$Q \ = \ \int \mathcal{N}_{d,v} \, R(dv)$$

with the limiting distribution R in (A2).

Corollary 2.2. *The random probability measure $\Gamma^\top P$ converges weakly in probability to the standard Gaussian distribution $\mathcal{N}_{d,1}$ if, and only if, the following condition is satisfied:*

(B) *For independent random vectors $X = X^{(q)}, \tilde{X} = \tilde{X}^{(q)}$ with distribution P,*

$$\|X\|^2/q \ \to_p \ 1 \quad \text{and} \quad X^\top \tilde{X}/q \ \to_p \ 0 \quad \text{as } q \to \infty. \qquad \square$$

The implication "(A2) $\Longrightarrow$ (A1)" in Theorem 2.1 as well as sufficiency of condition (B) in Corollary 2.2 are due to [4] (see their Theorem 1.1 and Proposition 4.2). They considered only (deterministic) empirical distributions P, but the extension to arbitrary distributions P is straightforward; see also Section 3.

It should be pointed out here that neither Theorem 2.1 nor Corollary 2.2 are just a consequence of Poincaré's [11] Lemma, although the latter is somehow at the heart of the proof. Poincaré showed that if $U_q = (U_{q,i})_{i=1}^q$ is uniformly distributed on the unit sphere in $\mathbb{R}^q$, then the Lebesgue density of $q^{1/2}U_{q,1}$ converges uniformly to the standard Gaussian density on $\mathbb{R}$. Translated into the present setting, one can show that for a fixed vector $x = x^{(q)} \in \mathbb{R}^q \setminus \{0\}$, the Lebesgue density of the random vector $\Gamma^\top x$ converges uniformly to the Lebesgue density of $\mathcal{N}_{d,v}$ as $q \to \infty$ and $\|x\|^2/q \to v > 0$.

Example 2.3. Condition (A2) is not a very restrictive requirement. For instance, suppose that $X = U(\mu_k + \sigma_k Z_k)_{k=1}^q$, where $(Z_k)_{k\geq 1}$ is a sequence of independent, identically distributed random variables with mean zero and variance one, while $U = U^{(q)}$ is an orthogonal matrix in $\mathbb{R}^{q\times q}$ and $\mu = \mu^{(q)} \in \mathbb{R}^q$, $\sigma = \sigma^{(q)} \in [0,\infty)^q$. Then condition (A2) is satisfied if, and only if,

$$\textbf{(A3)} \qquad \|\mu\|^2/q \ \to \ 0, \quad \|\sigma\|^2/q \ \to \ v \geq 0 \quad \text{and} \quad \max_{1\leq k\leq q} \sigma_k^2/q \ \to \ 0$$

as $q \to \infty$; see Section 4. Here $R = \delta_v$ and $Q = \mathcal{N}_{d,v}$.

Example 2.4. Suppose that $X \sim P^{(q)}$ has independent, identically distributed components such that

$$\mathbb{P}(X_i = \sqrt{q}) \ = \ 1 - \mathbb{P}(X_i = 0) \ = \ \pi_q,$$

where

$$\lim_{q\to\infty} q\pi_q \ = \ \lambda > 0.$$

Then $\mathcal{L}(\|X\|^2/q) = \mathrm{Bin}(q, \pi_q) \to_w \mathrm{Poiss}(\lambda)$ and $\mathcal{L}(X^\top \tilde{X}/q) = \mathrm{Bin}(q, \pi_q^2) \to_w \delta_0$ as $q \to \infty$. Hence (A2) is satisfied with $R = \mathrm{Poiss}(\lambda)$.

3. Empirical Distributions

From P to $\widehat{P}$. If the distributions $P = P^{(q)}$ satisfy conditions (A1-2), then the empirical distributions $\widehat{P} = \widehat{P}^{(q,n)}$ satisfy these conditions with high probability as $\min(q,n) \to \infty$. Precisely, one can easily deduce from condition (A2) that

$$D_{\mathrm{BL}}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{\|X_i\|^2/q}, R\Big) \ \to_p \ 0$$

and

$$\frac{1}{n^2}\sum_{i,j=1}^n \min\big\{|X_i^\top X_j/q|, 1\big\} \ \to_p \ 0$$

as $\min(q,n) \to \infty$. Thus Theorem 2.1 implies that

$$\Gamma^\top \widehat{P} = \frac{1}{n}\sum_{i=1}^n \delta_{\Gamma^\top X_i} \ \to_{w,p} \ \int \mathcal{N}_{d,v}\, R(dv)$$

as both q and n tend to infinity, where the random projector Γ and the empirical distribution $\widehat{P}$ are assumed to be stochastically independent.

Comparing P and $\widehat{P}$, part 1. In some sense Theorem 2.1 is a negative, though mathematically elegant result. It warns us against hasty conclusions about high-dimensional data sets after examining a couple of low-dimensional projections. In particular, one should not believe in multivariate normality only because several projections of the data "look normal". On the other hand, even small differences between different low-dimensional projections of $\widehat{P}$ may be intriguing. Therefore we study the relationship between projections of the empirical distribution $\widehat{P}$ and corresponding projections of P in more detail.

In particular, we are interested in the halfspace norm

$$\|\Gamma^\top \widehat{P} - \Gamma^\top P\|_{\mathrm{KS}} \ := \sup_{\text{closed halfspaces } H\subset\mathbb{R}^d} |\Gamma^\top \widehat{P}(H) - \Gamma^\top P(H)|$$

of $\Gamma^\top \widehat{P} - \Gamma^\top P$. In case of $d = 1$ this is the usual Kolmogorov-Smirnov norm of $\Gamma^\top \widehat{P} - \Gamma^\top P$. In what follows we use several well-known results from empirical process theory. Instead of citing original papers in various places we simply refer to the excellent monographs of [12] and [14]. It is known that

$$\mathbb{E} \sup_{\gamma\in\mathbb{R}^{q\times d}} \|\gamma^\top \widehat{P} - \gamma^\top P\|_{\mathrm{KS}} \ \le \ C\sqrt{q/n} \tag{1}$$

for some universal constant C. For the latter supremum is just the halfspace norm of $\widehat{P} - P$, and generally the set of closed halfspaces in $\mathbb{R}^k$ is a Vapnik-Cervonenkis class with Vapnik-Cervonenkis index $k + 1$. Inequality (1) does not capture the *typical* deviation between d-dimensional projections of $\widehat{P}$ and P. In fact,

$$\sup_{\gamma\in\mathbb{R}^{q\times d}} \mathbb{E}\, \|\gamma^\top \widehat{P} - \gamma^\top P\|_{\mathrm{KS}} \ \le \ C\sqrt{d/n},$$

which implies that

$$\mathbb{E}\, \|\Gamma^\top \widehat{P} - \Gamma^\top P\|_{\mathrm{KS}} \ \le \ C\sqrt{d/n}. \tag{2}$$

Our next result implies the limiting distribution of $\sqrt{n}\|\Gamma^\top \widehat{P} - \Gamma^\top P\|_{\mathrm{KS}}$ under conditions (A1-2). More generally, let $\mathcal{H}$ be a class of measurable functions from $\mathbb{R}^d$ into $[-1,1]$. Any finite signed measure M on $\mathbb{R}^d$ defines an element $h \mapsto M(h) := \int h\, dM$ of the space $\ell_\infty(\mathcal{H})$ of all bounded functions on $\mathcal{H}$ equipped with supremum norm $\|z\|_{\mathcal{H}} := \sup_{h\in\mathcal{H}} |z(h)|$. We shall impose the following three conditions on the class $\mathcal{H}$ and the distribution $Q = \int \mathcal{N}_{d,v}\, R(dv)$:

(C1) There exists a countable subset $\mathcal{H}_o$ of $\mathcal{H}$ auch that each $h \in \mathcal{H}$ can be represented as pointwise limit of some sequence in $\mathcal{H}_o$.

(C2) The set $\mathcal{H}$ satisfies the uniform entropy condition

$$\int_0^1 \sqrt{\log N(u,\mathcal{H})}\, du \ < \ \infty.$$

Here $N(u, \mathcal{H})$ is the supremum of $N(u, \mathcal{H}, \tilde{Q})$ over all probability measures $\tilde{Q}$ on $\mathbb{R}^d$, and $N(u, \mathcal{H}, \tilde{Q})$ is the smallest number m such that $\mathcal{H}$ can be covered with m balls having radius u with respect to the pseudodistance

$$\rho_{\tilde{Q}}(g, h) \ := \ \sqrt{\tilde{Q}((g-h)^2)}.$$

(C3) For any sequence $(Q_k)_k$ of probability measures converging weakly to Q,

$$\|Q_k - Q\|_{\mathcal{H}} \ \to \ 0 \quad \text{as } k \to \infty.$$

Condition (C1) ensures that random elements such as $\|\Gamma^\top \widehat{P} - \Gamma^\top P\|_{\mathcal{H}}$ are measurable. An example for conditions (C1-2) is the set $\mathcal{H}$ of (indicators of) closed halfspaces in $\mathbb{R}^d$. Then condition (C3) is a consequence of general results by [2], provided that $Q(\{0\}) = 0$, i.e. $R(\{0\}) = 0$.

A particular consequence of (C2) is existence of a centered Gaussian process B_Q, a so-called Q-bridge, having uniformly continuous sample paths with respect to ρ_Q and covariances

$$\mathbb{E}\big(B_Q(g)B_Q(h)\big) \ = \ Q(gh) - Q(g)Q(h),$$

which can be proved via a Chaining argument.

Theorem 3.1. *Suppose that the sequence $(P^{(q)})_{q \ge d}$ satisfies conditions (A1-2) of Theorem 2.1, and suppose that $\mathcal{H}$ fulfills conditions (C1-3). Then*

$$B^{(q,n)} \ := \ \Big(n^{1/2}\big(\Gamma^\top \widehat{P} - \Gamma^\top P\big)(h)\Big)_{h \in \mathcal{H}}$$

converges in distribution in $\ell_\infty(\mathcal{H})$ to B_Q as $\min(q, n) \to \infty$.

Comparing P and $\widehat{P}$, part 2. Theorem 3.1 takes into account the randomness in both the data (i.e. $\widehat{P}$) and the projection matrix Γ. However, exploratory projection pursuit means considering *several* projections of *one* data set. Thus we consider independent copies $\Gamma_\ell = \Gamma_\ell^{(q)}$, $\ell \ge 1$, of Γ which are also independent from $\widehat{P}$. With these projection matrices we define

$$B_\ell^{(q,n)} \ := \ \Big(n^{1/2}\big(\Gamma_\ell^\top \widehat{P} - \Gamma_\ell^\top P\big)(h)\Big)_{h \in \mathcal{H}}$$

and study the distribution of

$$\boldsymbol{B}^{(q,n)} \ := \ \big(B_\ell^{(q,n)}(h)\big)_{(\ell,h) \in \Lambda \times \mathcal{H}}$$

for $\Lambda := \{1, \ldots, L\}$ with an arbitrary fixed integer $L \ge 1$.

Subsequently a particular decomposition of the Q-Brigde B_Q will be used:

$$B_Q \ = \ B_Q' + B_Q''$$

with stochastically independent and centered Gaussian processes B_Q', B_Q'' on $\mathcal{H}$, where

$$\begin{aligned}
\mathbb{E}\big(B_Q'(g)B_Q'(h)\big) \ &= \ Q(gh) - \int \mathcal{N}_{d,v}(g)\,\mathcal{N}_{d,v}(h)\,R(dv) \\
&= \ \int \big(\mathcal{N}_{d,v}(gh) - \mathcal{N}_{d,v}(g)\mathcal{N}_{d,v}(h)\big)\,R(dv) \\
\mathbb{E}\big(B_Q''(g)B_Q''(h)\big) \ &= \ \int \mathcal{N}_{d,v}(g)\mathcal{N}_{d,v}(h)\,R(dv) - Q(g)Q(h).
\end{aligned}$$

By means of Anderson's Lemma (cf. [1]) or a further application of Chaining one can show that both B'_Q and B''_Q admit versions with uniformly continuous sample paths.

Theorem 3.2. *Suppose that the conditions of Theorem 3.1 are satisfied. Further, let* $B'_{Q,1}, B'_{Q,2}, B'_{Q,3}, \ldots$ *be independent copies of* B'_Q *and independent from* B''_Q. *Then for any fixed integer* $L \geq 1$, *the process* $\boldsymbol{B}^{(q,n)} = \bigl(B^{(q,n)}_\ell(h)\bigr)_{(\ell,h)\in\Lambda\times\mathcal{H}}$ *converges in distribution in* $\ell_\infty(\Lambda\times\mathcal{H})$ *to*

$$\boldsymbol{B} \ := \ \bigl(B'_{Q,\ell}(h) + B''_Q(h)\bigr)_{(\ell,h)\in\Lambda\times\mathcal{H}}$$

as $\min(q,n)\to\infty$.

Remark 3.3 (Understanding the decomposition $B_Q = B'_Q + B''_Q$ heuristically). Note that $B^{(q,n)}(h) = \sqrt{n}\int h(\Gamma^\top x)\,(\widehat{P}-P)(dx)$. Thus

$$\begin{aligned}\mathbb{E}\bigl(B^{(q,n)}(h)\,\big|\,\widehat{P}\bigr) &= \sqrt{n}\int \mathbb{E}\,h(\Gamma^\top x)\,(\widehat{P}-P)(dx)\\ &= \sqrt{n}\int \tilde{\mathcal{N}}_{d,q,\|x\|}(h)\,(\widehat{P}-P)(dx)\end{aligned}$$

with $\tilde{\mathcal{N}}_{d,q,\|x\|} := \mathcal{L}(\Gamma^\top x)$. Here we utilize orthogonal invariance of $\mathcal{L}(\Gamma)$. Consequently, $\mathbb{E}(B^{(q,n)}\,|\,\widehat{P})$ is a standardized empirical process indexed by the special functions $x \mapsto \tilde{\mathcal{N}}_{d,q,\|x\|}(h)$, $h\in\mathcal{H}$, and

$$\begin{aligned}&\mathbb{E}\Bigl(\mathbb{E}\bigl(B^{(q,n)}(g)\,\big|\,\widehat{P}\bigr)\,\mathbb{E}\bigl(B^{(q,n)}(h)\,\big|\,\widehat{P}\bigr)\Bigr)\\ &\quad= \int \tilde{\mathcal{N}}_{d,q,\|x\|}(g)\tilde{\mathcal{N}}_{d,q,\|x\|}(h)\,P(dx) - \int \tilde{\mathcal{N}}_{d,q,\|x\|}(g)\,P(dx)\int \tilde{\mathcal{N}}_{d,q,\|x\|}(h)\,P(dx).\end{aligned}$$

Since $\tilde{\mathcal{N}}_{d,q,\|x\|}$ is close to $\mathcal{N}_{d,\|x\|^2/q}$ and $\mathcal{L}(\|X\|^2/q)$ is close to R for large q, the latter covariance is close to

$$\int \mathcal{N}_{d,v}(g)\mathcal{N}_{d,v}(h)\,R(dv) - \int \mathcal{N}_{d,v}(g)\,R(dv)\int \mathcal{N}_{d,v}(h)\,R(dv) \ = \ \mathbb{E}\bigl(B''_Q(g)B''_Q(h)\bigr).$$

Example 3.4. Suppose that $d=1$, and let $\mathcal{H}$ consist of all indicator functions $1_{(-\infty,t]}$, $t\in\mathbb{R}$. Then Theorems 3.1 and 3.2 are applicable whenever $R(\{0\})=0$. Writing $M(t)$ instead of $M(1_{(-\infty,t]})$, the covariance functions of B_Q, B'_Q and B''_Q are given by

$$\begin{aligned}\mathbb{E}\bigl(B_Q(s)B_Q(t)\bigr) &= Q(\min\{s,t\}) - Q(s)Q(t),\\ \mathbb{E}\bigl(B'_Q(s)B'_Q(t)\bigr) &= Q(\min\{s,t\}) - \int \Phi(v^{-1/2}s)\Phi(v^{-1/2}t)\,R(dv),\\ \mathbb{E}\bigl(B''_Q(s)B''_Q(t)\bigr) &= \int \Phi(v^{-1/2}s)\Phi(v^{-1/2}t)\,R(dv) - Q(s)Q(t)\end{aligned}$$

for $s,t\in\mathbb{R}$, where $Q(u) = \int\Phi(v^{-1/2}u)\,R(dv)$, and Φ denotes the standard Gaussian distribution function.

Remark 3.5 (Conservative inference). Under conditions (A1-2) and (C1-3), *pretending* the empirical processes $B^{(q,n)}_\ell$, $1\leq\ell\leq L$, to be independent and identically distributed leads typically to conservative procedures. Precisely, let U be an open subset of $\ell_\infty(\mathcal{H})$. For instance let $U = \bigl\{b \in \ell_\infty(\mathcal{H}) : \|b\|_{\mathcal{H}} < \kappa\bigr\}$ for some constant $\kappa > 0$. Then it follows from Theorem 3.2 that

$$\liminf_{\min(q,n)\to\infty} \mathbb{P}\bigl(B_\ell^{(q,n)} \in U \text{ for } 1 \le \ell \le L\bigr) \ \ge\ \mathbb{P}(B_Q \in U)^L.$$

This may be verified as follows: By Theorem 3.2 and the Portmanteau Theorem, the limes inferior on the left hand side is not smaller than

$$\begin{aligned} \mathbb{P}\bigl(B'_{Q,\ell} + B''_Q \in U \text{ for } 1 \le \ell \le L\bigr) &= \mathbb{E}\,\mathbb{P}\bigl(B'_{Q,\ell} + B''_Q \in U \text{ for } 1 \le \ell \le L \,\big|\, B''_Q\bigr) \\ &= \mathbb{E}\Bigl(\mathbb{P}\bigl(B'_Q + B''_Q \in U \,\big|\, B''_Q\bigr)^L\Bigr), \end{aligned}$$

and by Jensen's inequality the latter expression is not smaller than

$$\Bigl(\mathbb{E}\,\mathbb{P}\bigl(B'_Q + B''_Q \in U \,\big|\, B''_Q\bigr)\Bigr)^L \ =\ \mathbb{P}(B'_Q + B''_Q \in U)^L \ =\ \mathbb{P}(B_Q \in U)^L.$$

If (A.1-2) is strengthened to (B) and $\mathbb{P}(B_Q \in \partial U) = 0$, then the previous arguments lead to

$$\left.\begin{array}{l} \lim\limits_{\min(q,n)\to\infty} \mathbb{P}\bigl(B_\ell^{(q,n)} \in U \text{ for } 1 \le \ell \le L\bigr) \\ \lim\limits_{\min(q,n)\to\infty} \mathbb{P}\bigl(B_\ell^{(q,n)} \in \overline{U} \text{ for } 1 \le \ell \le L\bigr) \end{array}\right\} \ =\ \mathbb{P}(B_Q \in U)^L,$$

because $B''_Q \equiv 0$ almost surely.

Remark 3.6 (The conditional point of view). Considering several projections of one data set means that we are interested in the *conditional* distribution of $n^{1/2}(\Gamma^\top \widehat{P} - \Gamma^\top P)$, *given* $\widehat{P}$. Indeed one may interpret Theorem 3.2 in the sense that for large q and n,

$$\mathcal{L}\bigl(B^{(q,n)} \,\big|\, \widehat{P}\bigr) \ \approx\ \mathcal{L}\bigl(B'_Q + B''_Q \,\big|\, B''_Q\bigr).$$

In case of the stronger condition (B) in Corollary 2.2, $B''_Q \equiv 0$, and

$$\mathcal{L}\bigl(B^{(q,n)} \,\big|\, \widehat{P}\bigr) \ \approx\ \mathcal{L}(B_Q).$$

Here are precise statements:

Corollary 3.7. *Suppose that the conditions of Theorem 3.1 are satisified. Let F be any bounded and continuous functional on $\ell_\infty(\mathcal{H})$ such that $F(B^{(q,n)})$ is measurable for all $q \ge d$ and $n \ge 1$. Then*

$$\mathbb{E}\bigl(F(B^{(q,n)}) \,\big|\, \widehat{P}\bigr) \ \to_{\mathcal{L}}\ \mathbb{E}\bigl(F(B'_Q + B''_Q) \,\big|\, B''_Q\bigr)$$

as $\min(q,n) \to \infty$. *In case of a degenerate distribution R,*

$$\mathbb{E}\bigl(F(B^{(q,n)}) \,\big|\, \widehat{P}\bigr) \ \to_p\ \mathbb{E}\,F(B_Q)$$

as $\min(q,n) \to \infty$.

4. Proofs

4.1. Hoeffding's (1952) trick

In connection with randomization tests, [7] observed that weak convergence of conditional distributions of test statistics is equivalent to the weak convergence of the *unconditional* distribution of suitable statistics in $\mathbb{R}^2$. His result can be extended straightforwardly as follows.

Lemma 4.1 (Hoeffding)**.** *For $k \ge 1$ let $X_k, \tilde{X}_k \in \mathbb{X}_k$ and $G_k \in \mathbb{G}_k$ be independent random variables, where $X_k, \tilde{X}_k$ are identically distributed. Further let m_k be some measurable mapping from $\mathbb{X}_k \times \mathbb{G}_k$ into the separable metric space $(\mathbb{M}, \rho)$, and let Q be a fixed Borel probability measure on $\mathbb{M}$. Then, as $k \to \infty$, the following two assertions are equivalent:*

(D1) $$\mathcal{L}\big(m_k(X_k, G_k) \,\big|\, G_k\big) \ \to_{w,p}\ Q.$$

(D2) $$\mathcal{L}\big(m_k(X_k, G_k), m_k(\tilde{X}_k, G_k)\big) \ \to_w\ Q \otimes Q.$$

Applications of this equivalence with non-Euclidean spaces $\mathbb{M}$ are presented by [13]. We shall utilize Lemma 4.1 in order to prove Theorem 2.1.

Proof of Lemma 4.1. Define $Y_k := m_k(X_k, G_k)$ and $\tilde{Y}_k := m_k(\tilde{X}_k, G_k)$. Suppose first that (D2) ist true, i.e. $\mathcal{L}(Y_k, \tilde{Y}_k) \to_w Q \otimes Q$. Then for any $f \in \mathcal{C}_b(\mathbb{M})$,

$$\begin{aligned}
&\mathbb{E}\Big(\big(\mathbb{E}(f(Y_k) \,|\, G_k) - Q(f)\big)^2\Big) \\
&\quad= \ \mathbb{E}\big(\mathbb{E}(f(Y_k) \,|\, G_k)^2\big) - 2Q(f)\ \mathbb{E}\,\mathbb{E}(f(Y_k) \,|\, G_k) + Q(f)^2 \\
&\quad= \ \mathbb{E}\,\mathbb{E}\big(f(Y_k) f(\tilde{Y}_k) \,\big|\, G_k\big) - 2Q(f)\ \mathbb{E}\,\mathbb{E}(f(Y_k) \,|\, G_k) + Q(f)^2 \\
&\quad= \ \mathbb{E}\big(f(Y_k) f(\tilde{Y}_k)\big) - 2Q(f)\,\mathbb{E}\, f(Y_k) + Q(f)^2 \\
&\quad\to \ \int f(y) f(\tilde{y})\, Q(dy) Q(d\tilde{y}) - Q(f)^2 \\
&\quad= \ 0.
\end{aligned}$$

Thus $\mathcal{L}(Y_k \,|\, G_k) \to_{w,p} Q$.

On the other hand, suppose that (D1) is satisfied, i.e. $\mathcal{L}(Y_k \,|\, G_k) \to_{w,p} Q$. Then for arbitrary $f, g \in \mathcal{C}_b(\mathbb{M})$,

$$\begin{aligned}
\mathbb{E}\big(f(Y_k) g(\tilde{Y}_k)\big) \ &= \ \mathbb{E}\,\mathbb{E}\big(f(Y_k) g(\tilde{Y}_k) \,\big|\, G_k\big) \\
&= \ \mathbb{E}\big(\mathbb{E}(f(Y_k) \,|\, G_k)\ \mathbb{E}(f(\tilde{Y}_k) \,|\, G_k)\big) \\
&\to \ Q(f) Q(g),
\end{aligned}$$

because $\mathbb{E}(h(Y_k) \,|\, G_k) \ \to_p\ \int h\, dQ$ and $\big|\mathbb{E}(h(Y_k) \,|\, G_k)\big| \ \le\ \|h\|_\infty \ <\ \infty$ for each $h \in \mathcal{C}_b(\mathbb{M})$. Thus we know that $\mathbb{E}\, F(Y_k, \tilde{Y}_k) \to \int F\, dQ \otimes Q$ for arbitrary functions $F(y, \tilde{y}) = f(y) g(\tilde{y})$ with $f, g \in \mathcal{C}_b(\mathbb{M})$. But this is known to be equivalent to weak convergence of $\mathcal{L}(Y_k, \tilde{Y}_k)$ to $Q \otimes Q$; see Chapter 1.4 of [14].

Here is an alternative argument: With $\widehat{Q}_k := \mathcal{L}(Y_k \,|\, G_k)$, Assumption (D1) is equivalent to $D_{\mathrm{BL}}(\widehat{Q}_k, Q) \to_p 0$. To prove that $\mathcal{L}(Y_k, \tilde{Y}_k) \to Q \otimes Q$, it suffices to show that $\mathbb{E}\big(F(Y_k, \tilde{Y}_k) \,\big|\, G_k\big) \to_p \int F\, dQ \otimes Q$ for any function $F : \mathbb{M} \times \mathbb{M} \to [-1, 1]$ such that $\big|F(y, \tilde{y}) - F(z, \tilde{z})\big| \le \rho(y, z) + \rho(\tilde{y}, \tilde{z})$ for arbitrary $y, \tilde{y}, z, \tilde{z} \in \mathbb{M}$. But this entails that $F(y, \cdot), F(\cdot, \tilde{y}) \in \mathcal{F}_{\mathrm{BL}}$ for arbitrary $y, \tilde{y} \in \mathbb{M}$. Consequently,

$$\begin{aligned}
&\Big|\mathbb{E}\big(F(Y_k, \tilde{Y}_k) \,\big|\, G_k\big) - \int F\, dQ \otimes Q\Big| \\
&\quad= \ \Big|\int F\, d\big(\widehat{Q}_k \otimes \widehat{Q}_k - Q \otimes Q\big)\Big| \\
&\quad\le \ \int \Big|\int F(\cdot, \tilde{y})\, d\big(\widehat{Q}_k - Q\big)\Big|\, \widehat{Q}_k(d\tilde{y}) + \int \Big|\int F(y, \cdot)\, d\big(\widehat{Q}_k - Q\big)\Big|\, Q(dy) \\
&\quad\le \ 2 D_{\mathrm{BL}}(\widehat{Q}_k, Q).
\end{aligned}$$

□

4.2. Proofs for Section 2

That $\Gamma = \Gamma^{(q)}$ is "uniformly" distributed on the set of column-wise orthonormal matrices in $\mathbb{R}^{q\times d}$ means that $\mathcal{L}(U\Gamma) = \mathcal{L}(\Gamma)$ for any fixed orthonormal matrix $U \in \mathbb{R}^{q\times q}$. For existence and uniqueness of the latter distribution we refer to Chapters 1-2 of [6]. For the present purposes the following explicit construction of Γ described in Chapter 7 of [6] is sufficient. Let $Z = Z^{(q)} := (Z_1, Z_2, \ldots, Z_d)$ be a random matrix in $\mathbb{R}^{q\times d}$ with independent, standard Gaussian column vectors $Z_j \in \mathbb{R}^q$. Then

$$\Gamma \ := \ Z(Z^\top Z)^{-1/2}$$

has the desired distribution, and

$$\Gamma \ = \ q^{-1/2} Z \,(I + O_p(q^{-1/2})) \quad \text{as } q \to \infty. \tag{3}$$

This equality can be viewed as an extension of Poincaré's [11] Lemma.

Proof of Theorem 2.1. Let $\Gamma = \Gamma(Z)$ as above. Suppose that $Z = Z^{(q)}$, $X = X^{(q)}$ and $\tilde{X} = \tilde{X}^{(q)}$ are independent with $\mathcal{L}(X) = \mathcal{L}(\tilde{X}) = P$, and let $Y, \tilde{Y}$ be two independent random vectors in $\mathbb{R}^d$ with distribution Q. According to Lemma 4.1, condition (A1) is equivalent to

$$\begin{pmatrix}\Gamma^\top X\\ \Gamma^\top \tilde{X}\end{pmatrix} \ \to_{\mathcal{L}} \ \begin{pmatrix}Y\\ \tilde{Y}\end{pmatrix}. \tag{$\mathbf{A1'}$}$$

Because of equation (3) this can be rephrased as

$$\begin{pmatrix}Y^{(q)}\\ \tilde{Y}^{(q)}\end{pmatrix} := \begin{pmatrix}q^{-1/2} Z^\top X\\ q^{-1/2} Z^\top \tilde{X}\end{pmatrix} \ \to_{\mathcal{L}} \ \begin{pmatrix}Y\\ \tilde{Y}\end{pmatrix}. \tag{$\mathbf{A1''}$}$$

Now we prove equivalence of (A1") and (A2) starting from the observation that

$$\mathcal{L}\left(\begin{pmatrix}Y^{(q)}\\ \tilde{Y}^{(q)}\end{pmatrix}\right) \ = \ \mathbb{E}\,\mathcal{L}\left(\begin{pmatrix}Y^{(q)}\\ \tilde{Y}^{(q)}\end{pmatrix} \,\Big|\, X, \tilde{X}\right) \ = \ \mathbb{E}\,\mathcal{N}_{2d}(0, \Sigma^{(q)}),$$

where

$$\Sigma^{(q)} \ := \ \begin{pmatrix} q^{-1}\|X\|^2\, I_d & q^{-1} X^\top \tilde{X}\, I_d\\ q^{-1} X^\top \tilde{X}\, I_d & q^{-1}\|\tilde{X}\|^2\, I_d\end{pmatrix} \ \in \ \mathbb{R}^{2d\times 2d}.$$

Suppose that condition (A2) holds. Then $\Sigma^{(q)}$ converges in distribution to a random diagonal matrix

$$\Sigma \ := \ \begin{pmatrix} S^2\, I_d & 0\\ 0 & \tilde{S}^2\, I_d\end{pmatrix}$$

with independent random variables $S^2, \tilde{S}^2$ having distribution R. Clearly this implies that

$$\mathbb{E}\,\mathcal{N}_{2d}(0, \Sigma^{(q)}) \ \to_w \ \mathbb{E}\,\mathcal{N}_{2d}(0, \Sigma) \ = \ \mathcal{L}\left(\begin{pmatrix}Y\\ \tilde{Y}\end{pmatrix}\right)$$

with $Q = \mathbb{E}\,\mathcal{N}_d(0, S^2 I_d)$. Hence (A1") holds.

On the other hand, suppose that (A1") holds. For any $t = (t_1^\top, t_2^\top)^\top \in \mathbb{R}^{2d}$, the Fourier transform of $\mathcal{L}\big((Y^{(q)\top}, \tilde{Y}^{(q)\top})^\top\big)$ at t equals

$$\mathbb{E}\,\exp\big(\boldsymbol{i}\,(t_1^\top Y^{(q)} + t_2^\top \tilde{Y}^{(q)})\big) \ = \ \mathbb{E}\,\exp(-t^\top \Sigma^{(q)} t/2) \ = \ H^{(q)}(a(t)),$$

where $\boldsymbol{i}$ stands for $\sqrt{-1}$, $a(t) := \bigl(\|t_1\|^2/2, \|t_2\|^2/2, t_1^\top t_2\bigr)^\top \in \mathbb{R}^3$, and

$$H^{(q)}(a) \ := \ \mathbb{E} \exp\bigl(-a_1\|X\|^2/q - a_2\|\tilde{X}\|^2/q - a_3 X^\top \tilde{X}/q\bigr)$$

denotes the Laplace transform of $\mathcal{L}\bigl(\bigl(\|X\|^2/q, \|\tilde{X}\|^2/q, X^\top \tilde{X}/q\bigr)^\top\bigr)$ at $a \in \mathbb{R}^3$. By assumption, the Fourier transform at t converges to

$$\mathbb{E} \exp(\boldsymbol{i}\, t_1^\top Y)\ \mathbb{E} \exp(\boldsymbol{i}\, t_2^\top Y).$$

Setting $t_2 = 0$ and varying t_1 shows that the Laplace transform of $\mathcal{L}(\|X\|^2/q)$ converges pointwise on $[0,\infty)$ to a continuous function. Hence $\|X\|^2/q$ converges in distribution to some random variable $S^2 \geq 0$, and $Q = \mathbb{E}\mathcal{N}_{d,S^2}$. Therefore, if $\tilde{S}^2$ denotes an independent copy of S^2, we know that $H^{(q)}(a(t))$ converges to

$$\mathbb{E} \exp(-a_1(t)S^2)\, \mathbb{E} \exp(-a_2(t)S^2) \ = \ \mathbb{E} \exp\bigl(-a_1(t)S^2 - a_2(t)\tilde{S}^2 - a_3(t)\cdot 0\bigr).$$

A problem at this point is that for dimension $d = 1$ the set $\{a(t) : t \in \mathbb{R}^{2d}\} \subset \mathbb{R}^3$ has empty interior. Thus we cannot apply the standard argument about weak convergence and convergence of Laplace transforms. However, letting $t_2 = \pm t_1$ with $\|t_1\|^2/2 = 1$, one may conclude that

$$\begin{aligned}
0 \ &= \ \lim_{q\to\infty} \bigl(H^{(q)}(1,1,2) + H^{(q)}(1,1,-2) - 2H^{(q)}(1,0,0)^2\bigr) \\
&= \ \lim_{q\to\infty} \bigl(H^{(q)}(1,1,2) + H^{(q)}(1,1,-2) - 2\ \mathbb{E}\exp(-\|X\|^2/q - \|\tilde{X}\|^2/q)\bigr) \\
&= \ 2 \lim_{q\to\infty} \mathbb{E}\Bigl(\exp\bigl(-\|X\|^2/q - \|\tilde{X}\|^2/q\bigr)\bigl(\cosh(2X^\top\tilde{X}/q) - 1\bigr)\Bigr).
\end{aligned}$$

But for arbitrary small $\epsilon > 0$ and large $r > 0$,

$$\begin{aligned}
&\mathbb{E}\Bigl(\exp\bigl(-\|X\|^2/q - \|\tilde{X}\|^2/q\bigr)\bigl(\cosh(2X^\top\tilde{X}/q) - 1\bigr)\Bigr) \\
&\quad \geq \ \exp(-2r)(\cosh(2\epsilon) - 1)\ \mathbb{P}\bigl(\|X\|^2/q < r, \|\tilde{X}\|^2/q < r, |X^\top\tilde{X}/q| \geq \epsilon\bigr) \\
&\quad \geq \ \exp(-2r)(\cosh(2\epsilon) - 1)\ \Bigl(\mathbb{P}\bigl(|X^\top\tilde{X}/q| \geq \epsilon\bigr) - 2\,\mathbb{P}(\|X\|^2/q \geq r)\Bigr) \\
&\quad \geq \ \exp(-2r)(\cosh(2\epsilon) - 1)\ \Bigl(\mathbb{P}\bigl(|X^\top\tilde{X}/q| \geq \epsilon\bigr) - 2\,\mathbb{P}(S^2 \geq r) + o(1)\Bigr).
\end{aligned}$$

Hence

$$\limsup_{q\to\infty}\ \mathbb{P}\bigl(|X^\top\tilde{X}/q| \geq \epsilon\bigr) \ \leq \ 2\,\mathbb{P}(S^2 \geq r).$$

Letting $r \to \infty$ shows that $X^\top\tilde{X}/q \to_p 0$. □

Proof of equivalence of (A2) and (A3). Proving that (A3) implies (A2) is elementary. In order to show that (A2) implies (A3) note first that conditions (A2) for the distributions $P^{(q)}$ imply the same conditions for the symmetrized distributions

$$P_o = P_o^{(q)} \ := \ \mathcal{L}(X - \tilde{X}) \ = \ \mathcal{L}\Bigl(\bigl(\sigma_k(Z_k - Z_{q+k})\bigr)_{1\leq k\leq q}\Bigr).$$

Condition (A2) for these distributions reads as follows.

$$\mathcal{L}\Bigl(\sum_{k=1}^q (Z_k - Z_{q+k})^2\sigma_k^2/q\Bigr) \quad \to_w \quad R_o = R \star R \quad \text{and} \tag{4}$$

$$\sum_{k=1}^q (Z_k - Z_{q+k})(Z_{2q+k} - Z_{3q+k})\sigma_k^2/q \quad \to_p \quad 0. \tag{5}$$

The factors $(Z_k - Z_{q+k})(Z_{2q+k} - Z_{3q+k})$, $1 \le k \le q$, in (5) are independent, identically and symmetrically distributed. By conditioning on any one of these factors one can deduce from (5) that $\max_{1 \le k \le q} \sigma_k^2/q \to 0$. But then

$$\sum_{k=1}^{q} \sigma_k^2 (Z_k - Z_{q+k})^2/q \ = \ 2\|\sigma\|^2/q + o_p(1 + \|\sigma\|^2/q),$$

and one can deduce from (4) that $\|\sigma\|^2/q$ converges to some fixed number v; in particular, $R = \delta_v$. Now we return to the original distributions P. Here the second half of (A2) means that

$$\begin{aligned} &\sum_{k=1}^{k} (\mu_k + \sigma_k Z_k)(\mu_k + \sigma_k Z_{q+k})/q \\ &= \ \|\mu\|^2/q + \sum_{k=1}^{q} \mu_k \sigma_k (Z_k + Z_{q+k})/q + \sum_{k=1}^{q} \sigma_k^2 Z_k Z_{q+k}/q \\ &= \ o_p(1). \end{aligned}$$

Since

$$\begin{aligned} \mathbb{E}\Big(\Big(\sum_{k=1}^{q} \mu_k \sigma_k (Z_k + Z_{q+k})/q \Big)^2 \Big) \ &= \ \sum_{k=1}^{q} \mu_k^2 \sigma_k^2 / q^2 \ = \ o(\|\mu\|^2/q), \\ \mathbb{E}\Big(\Big(\sum_{k=1}^{q} \sigma_k^2 Z_k Z_{q+k}/q \Big)^2 \Big) \ &= \ \sum_{k=1}^{q} \sigma_k^4/q^2 \ \to \ 0, \end{aligned}$$

it follows that $\|\mu\|^2/q \to 0$. □

4.3. Proofs for Section 3

Since Theorem 3.1 is just Theorem 3.2 with $L = 1$, it suffices to verify the latter.

Proof of Theorem 3.2. It suffices to verify the following two claims:

(F1) As $q \to \infty$ and $n \to \infty$, the finite-dimensional marginal distributions of the process $\boldsymbol{B}^{(q,n)}$ converge to the corresponding finite-dimensional distributions of $\boldsymbol{B}$.
(F2) As $q \to \infty$, $n \to \infty$ and $\delta \downarrow 0$,

$$\max_{\ell \in \Lambda} \sup_{g,h \in \mathcal{H} : \rho_Q(g,h) < \delta} \Big| B_\ell^{(q,n)}(g) - B_\ell^{(q,n)}(h) \Big| \ \to_p \ 0.$$

The second condition, (F2), means that the processes $\boldsymbol{B}^{(q,n)}$ are asymptotically equicontinuous with respect to the pseudodistance

$$\boldsymbol{\rho}_Q\big((\ell, g), (m, h)\big) \ := \ 1\{\ell \ne m\} + \rho_Q(g,h)$$

on $\Lambda \times \mathcal{H}$.

In order to verify assertions (F1-2) we consider the conditional distribution of $\boldsymbol{B}^{(q,n)}$ given the random matrix

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{(q)} \ := \ (\Gamma_1, \Gamma_2, \ldots, \Gamma_L) \ \in \ \mathbb{R}^{q \times Ld}.$$

In fact, if we define

$$f_{\ell,h}(\boldsymbol{v}) \ := \ h(v_\ell) \quad \text{for } \boldsymbol{v} = (v_1^\top, \ldots, v_L^\top)^\top \in \mathbb{R}^{Ld},$$

then

$$B_\ell^{(q,n)}(h) \ = \ n^{1/2}(\boldsymbol{\Gamma}^\top \widehat{P} - \boldsymbol{\Gamma}^\top P)(f_{\ell,h}).$$

Thus $\mathcal{L}(\boldsymbol{B}^{(q,n)} \,|\, \boldsymbol{\Gamma})$ is essentially the distribution of an empirical process based on n independent random vectors with distribution $\boldsymbol{\Gamma}^\top P$ on $\mathbb{R}^{Ld}$ and indexed by the family $\tilde{\mathcal{H}} := \{f_{\ell,h} : \ell \in \Lambda, h \in \mathcal{H}\}$.

The multivariate version of Lindeberg's Central Limit Theorem entails that for large q and n, the finite-dimensional marginal distributions of $\boldsymbol{B}^{(q,n)}$, conditional on $\boldsymbol{\Gamma}$, can be approximated by the corresponding finite-dimensional distributions of a centered Gaussian process on $\Lambda \times \mathcal{H}$ with the same covariance function, namely,

$$\begin{aligned} \Sigma^{(q)}\bigl((\ell,g),(m,h)\bigr) \ &:= \ \mathrm{Cov}\bigl(B_\ell^{(q,n)}(g), B_m^{(q,n)}(h) \,\big|\, \boldsymbol{\Gamma}\bigr) \\ &= \ \boldsymbol{\Gamma}^\top P(f_{\ell,g} f_{m,h}) - \boldsymbol{\Gamma}^\top P(f_{\ell,g}) \boldsymbol{\Gamma}^\top P(f_{m,h}). \end{aligned}$$

It follows from equality (3) and the proof of Theorem 2.1 that

$$\boldsymbol{\Gamma}^\top P \ \to_{w,p} \ \boldsymbol{Q} \ := \ \int \mathcal{N}_{Ld,v}\, R(dv) \quad \text{as } q \to \infty,$$

and this should imply convergence of $\Sigma^{(q)}$ to some limiting function as well. It was shown by [2] that condition (C3) is equivalent to

$$\lim_{\delta \downarrow 0} \sup_{h \in \mathcal{H}} Q\Bigl\{ y \in \mathbb{R}^d : \sup_{z : \|z - y\| < \delta} |h(z) - h(y)| > \epsilon \Bigr\} \ = \ 0 \quad \text{for any } \epsilon > 0. \tag{6}$$

Note that the d-dimensional marginal distributions of $\boldsymbol{Q}$ are just Q. Therefore one can easily deduce from (6) that for any fixed $\epsilon > 0$,

$$\lim_{\delta \downarrow 0} \sup_{f', f'' \in \tilde{\mathcal{H}} \cup \{1\}} \boldsymbol{Q}\Bigl\{ \boldsymbol{v} \in \mathbb{R}^{Ld} : \sup_{\boldsymbol{w} : \|\boldsymbol{w} - \boldsymbol{v}\| < \delta} |f'f''(\boldsymbol{w}) - f'f''(\boldsymbol{v})| > \epsilon \Bigr\} \ = \ 0.$$

Hence a second application of [2] shows that

$$\sup_{f', f'' \in \tilde{\mathcal{H}} \cup \{1\}} |\boldsymbol{\Gamma}^\top P(f'f'') - \boldsymbol{Q}(f'f'')| \ \to \ 0 \quad \text{as } q \to \infty, \tag{7}$$

because $\boldsymbol{\Gamma}^\top P \to_{w,p} \boldsymbol{Q}$. In particular, the conditional covariance function $\Sigma^{(q)}$ converges uniformly in probability to the covariance function Σ, where

$$\begin{aligned} \Sigma\bigl((\ell,g),(m,h)\bigr) \ &:= \ \boldsymbol{Q}(f_{\ell,g} f_{m,h}) - \boldsymbol{Q}(f_{\ell,g}) \boldsymbol{Q}(f_{m,h}) \\ &= \ \int \mathcal{N}_{Ld,v}(f_{\ell,g} f_{m,h})\, R(dv) - Q(g)Q(h) \\ &= \ \begin{cases} \displaystyle\int \mathcal{N}_{d,v}(gh)\, R(dv) - Q(g)Q(h) & \text{if } \ell = m, \\ \displaystyle\int \mathcal{N}_{d,v}(g) \mathcal{N}_{d,v}(h)\, R(dv) - Q(g)Q(h) & \text{if } \ell \neq m, \end{cases} \\ &= \ \mathrm{Cov}\bigl(B'_{Q,\ell}(g) + B''_Q(g), B'_{Q,m}(h) + B''_Q(h)\bigr) \end{aligned}$$

as $q \to \infty$. This proves assertion (F1).

As for assertion (F2), it is well-known from empirical process theory that conditions (C1-2) imply that for arbitrary fixed $\epsilon > 0$,

$$\max_{\ell \in \Lambda} \mathbb{P}\Big(\sup_{g,h \in \mathcal{H} : \rho_\ell^{(q)}(g,h) < \delta} \Big| B_\ell^{(q,n)}(g) - B_\ell^{(q,n)}(h) \Big| \geq \epsilon \,\Big|\, \mathbf{\Gamma} \Big) \ \to_p \ 0 \tag{8}$$

as $\min(q,n) \to \infty$ and $\delta \downarrow 0$. Here

$$\rho_\ell^{(q)}(g,h) \ := \ \sqrt{\mathbf{\Gamma}^\top P((f_{\ell,g} - f_{\ell,h})^2)} \ = \ \sqrt{\Gamma_\ell^\top P((g-h)^2)}.$$

But it follows from (7) that

$$\max_{\ell \in \Lambda} \sup_{g,h \in \mathcal{H}} |\rho_\ell^{(q)}(g,h)^2 - \rho_Q(g,h)^2| \ \to_p \ 0$$

as $q \to \infty$. Hence one may replace $\rho_\ell^{(q)}$ in (8) with ρ_Q and obtain assertion (F2). $\square$

Proof of Corollary 3.7. The main trick is to replace conditional expectations with suitable sample means. Note that conditional on $\widehat{P}$, the processes $B_1^{(q,n)}$, $B_2^{(q,n)}$, $B_3^{(q,n)}$, ... are independent copies of $B^{(q,n)}$. Likewise, conditional on B_Q'', the processes $B_{Q,1}' + B_Q'', B_{Q,2}' + B_Q'', B_{Q,3}' + B_Q'', \ldots$ are independent copies of $B_Q' + B_Q''$. Hence

$$\left.\begin{aligned} \mathbb{E}\Big|\mathbb{E}\big(F(B^{(q,n)}) \,\big|\, \widehat{P}\big) - L^{-1}\sum_{\ell=1}^L F(B_\ell^{(q,n)})\Big| \\ \mathbb{E}\Big|\mathbb{E}\big(F(B_Q' + B_Q'') \,\big|\, B_Q''\big) - L^{-1}\sum_{\ell=1}^L F(B_{Q,\ell}' + B_Q'')\Big| \end{aligned}\right\} \ \leq \ L^{-1/2}\|F\|_\infty$$

for any integer $L \geq 1$. Consequently it suffices to show that for any fixed $L \geq 1$, the random variable $L^{-1}\sum_{\ell=1}^L F(B_\ell^{(q,n)})$ converges in distribution to the random variable $L^{-1}\sum_{\ell=1}^L F(B_{Q,\ell}' + B_Q'')$ as $\min(q,n) \to \infty$. But this is a consequence of Theorem 3.2 and the Continuous Mapping Theorem, because

$$\boldsymbol{b} = \big(b_\ell(h)\big)_{(\ell,h) \in \Lambda \times \mathcal{H}} \ \mapsto \ L^{-1}\sum_{\ell=1}^L F(b_\ell)$$

defines a continuous mapping from $\ell_\infty(\Lambda \times \mathcal{H})$ to $\mathbb{R}$. $\square$

Acknowledgement. Part of this work is contained in the diploma thesis of Perla Zerial (1995, Univ. of Heidelberg). We are grateful to Jon Wellner, Aad van der Vaart and an anonymous referee for their interest in this work, stimulating discussions and pertinent questions.

References

[1] Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* ***6***, 170-176.

[2] Billingsley, P. and F. Topsoe (1967). Uniformity in weak convergence. *Z. Wahrschein. verw. Geb.* ***7***, 1-16.

[3] Buja, A., D. Cook and D.F. Swayne (1996). Interactive High-Dimensional Data Visualization. *J. Comp. Graph. Statist.* ***5***, 78-99.

[4] Diaconis, P. and D. Freedman (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* ***12***, 793-815.

[5] Eaton, M.L. (1981). On the projections of isotropic distributions. *Ann. Statist.* ***9***, 391-400.

[6] Eaton, M.L. (1989). *Group Invariance Applications in Statistics.* Regional Conf. Series Prob. Statist. **1**, IMS.

[7] Hoeffding, W. (1952). The large-sample power of tests based on random permutations. *Ann. Math. Statist.* ***23***, 169-192.

[8] Huber, P.J. (1985). Projection pursuit (with discussion). *Ann. Statist.* ***13***, 435-475.

[9] Meckes, E. (2009). Quantitative asymptotics of graphical projection pursuit. *Electron. Comm. Probab.* ***14***, 176-185.

[10] Meckes, E. (2011). Projections of probability distributions: A measure-theoretic Dvoretzky theorem. Preprint (arXiv:1102.3438)

[11] Poincaré, H. (1912). *Calcul des Probabilités.* Hermann, Paris.

[12] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

[13] Romano, J.P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* ***17***, 141-159.

[14] van der Vaart, A.W. and J.A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York.